基于因果效应的贝叶斯网络结构学习方法 *

安　宁，滕　越，杨矫云†，李　廉

(合肥工业大学 国家智慧养老国际科技合作基地, 合肥 230009)

摘　要：从数据中学习贝叶斯网络结构是一个 NP-hard 问题，提高网络结构学习算法精度是研究的重难点。基于 Judea Pearl 因果理论，提出了一种贝叶斯网络结构学习方法，提升了现有算法的准确率。利用改进的 Pearl 因果效应和 BDe 评分，学习网络节点优先次序，利用 K2 算法学习初始网络，并通过 BDe 评分反向调节、互信息和 BDe 评分删除边以修正学习结果。实验在贝叶斯网络标准数据集 ASIA、ALARM 上进行，在样本量为 2000~20000 的 20 组实验中，学习准确率较 MMHC 算法平均提升 16%，准确率标准差较 MMHC 算法平均缩小 17%。实验表明，基于因果效应的方法较 MMHC 算法有更好地性能。

关键词：贝叶斯网络；阿尔茨海默病；K2 算法；因果效应；BDe 评分；互信息

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2018.07.0348

Bayesian network structure learning method based on causal effect

An Ning, Teng Yue, Yang Jiaoyun†, Li Lian

(*National Smart Eldercare International S&T Cooperation Base*, *School of Computer & Information*, *Hefei University of Technology*, *Hefei* 230009, *China*)

Abstract: Learning bayesian networks from data is a NP-hard problem. Improving the accuracy of network structure learning algorithms is important, yet difficulty for researchers. The paper proposed a Bayesian network structure learning algorithm based on Judea Pearl's causal theory to improve the accuracy of existing algorithms. The algorithm made use of the improved causal effect and the BDe scoring-function to learn the nodes order. then applied K2 algorithm to achieve an initialized network. At last, it used the BDe scoring-function to adjust the network inversely, besides, it used mutual-information and BDe scores to prune the edges. The proposed method outperforms MMHC algorithm on the ASIA and ALARM data set, with 16% accurate improvement, and 17% standard deviation less on average. The results show that the method base on causal effect is batter then MMHC algorithm.

Key words: Bayesian network; Alzheimer's disease; K2 algorithm; Causal effect; BDe scoring-function; Mutual information

0　引言

贝叶斯网络是不确定知识表达的重要工具、是一些节点联合概率分布的有向无环图，它由有向无环图（directed acyclic graph）和条件概率表（conditional probability table）两部分组成。其中，有向无环图定性的表示了变量之间的独立关系，条件概率表则定量的表示了变量间的依赖程度。由于其图形可视化的特点，被广泛应用于生物医药领域[1,2]、预测领域[3,4]、分类[5,6]、因果推断[7,8]、视觉识别[9]、信息检索[10]等。

贝叶斯网的构建方法主要分为通过专家知识构建和通过数据构建。通过专家知识构建贝叶斯网络是一项非常繁琐且容易出错的工作，而从数据中学习得到贝叶斯网络是一个 NP-hard 问题[11]。因此，如何高效、高质量地从数据中学习贝叶斯网络成为了研究的重难点之一。近几十年来，涌现了很多结构学习方法，基于条件独立测试的算法[12,13]、基于评分搜索的算法[14~16]以及混合条件独立性测试和评分搜索的方法[17]。

基于评分函数的方法包含两个部分：a）用来衡量网络和数据拟合程度的评分函数,如 BIC（Bayesian information criterion）评分标准，它是在样本满足独立同分布的前提下，用对数似然度来度量网络结构与观测数据的拟合程度[18]、还有假设结构先验分布服从狄利克雷分布的 BDe 评分[15]、假设结构先验分布服从均匀分布的 CH 评分（K2 评分函数）[19]、基于数据编码总长

收稿日期：2018-07-10；**修回日期：**2018-08-24　　**基金项目：**安徽省重点研究与开发计划项目（1704e1002221）；国家自然计学基金青年基金资助项目（61502135）；国家高等学校学科创新引智计划资助项目（B14025）

作者简介：安宁（1971-），男，教授，博士，主要研究方向为空间数据管理、公共危机信息管理、移动健康、普适计算；滕越（1992-），男，硕士研究生，主要研究方向为机器学习、因果学习；杨矫云（1987-）男（通信作者），副教授，博士，主要研究方向为健康信息挖掘、因果理论（jiaoyun@hfut.edu.cn）；李廉（1951-），男，教授，博士，主要研究方向为因果理论、机器学习．

度度量的 MDL 评分（minimum description length）[14]等；b）寻找最高评分函数的方法。如：需要节点优先顺序的 K2 算法[21,22]、最大最小爬山算法[23]、更容易获得全局最优的粒子群算法[20]、爬山算法等。由于搜索空间是关于节点个数指数上升的，所以不能使用穷举法得到评分最高的网络。

基于条件独立性测试的方法虽然能较准确的学习贝叶斯网络结构，但只适用于离散数据，对于多维数据下的学习效率不理想。结合条件独立性测试和评分搜索的方法有更高的准确率。其中最为代表性的是 MMHC 方法。MMHC 算法分为第一阶段 MMPC 算法，和第二阶段的评分方法。MMPC 算法首先确定目标变量 *T* 的父子节点集 *CPC(T)*，以 *T* 变量的父子节点集为条件，计算网络中其他变量关于 *T* 变量的最小依赖程度 *Assoc(X,T|CPC(T))*。最小依赖程度取最大的变量加入到 *T* 的候选父子节点集。直到 *MaxAssoc(X,T|CPC(T))*为零进入 MMPC 算法第二阶段，通过两个准则删除第一阶段错误加入的候选父子节点，从而构建无方向的贝叶斯骨架。MMHC 算法第二阶段则是通过删除、增加、反向三个调节因子结合评分函数进行打分，取最高分值作为网络学习结果[23]。

K2 算法是一种可以明显缩小搜索空间的贪婪搜索算法，但需要提供节点优先顺序（在序列中，排在前的节点不可能是排在后的节点的子节点）。节点顺序的正确性，直接影响了网络的训练结果。近年来，有部分学者针对 K2 节点优先顺序做了相关研究[17]，从算法的时间复杂度和准确性来说，仍不能满足大数据下的应用。

基于因果效应的方法从 Judea Pearl 提出的因果理论入手[24]，通过定义的因果效应强度计算、无边网络 BDe 评分计算，得到节点优先顺序。结合 K2 算法和互信息学习贝叶斯网络。在标准数据库下的实验表明，该方法在准确率和标准差性能上远高于其他方法。

1　贝叶斯网络和因果推断

1.1　贝叶斯网络

贝叶斯网是一个带参数的有向无环图，可以用 $<G,\Theta>$ 来表示。其中：*G* 代表有向无环图，如图 1 所示。Θ 代表父节点对于子节点的条件概率表。*G* 是一个二元组 $\langle V,G\rangle$，其中 *V* 代表所有随机变量组成的点集，$V=\{X_1,\dots,X_n\}$，X_i 为图 G 中第 i 个节点，*E* 代表边组成的集合，根据图论的相关知识，边集用矩阵 A_{n*n} 来表示。若$X_i \rightarrow X_j$，则矩阵 $A_{i,j}=1$，否则 $A_{i,j}=0$。在有向无环图中，节点代表每一个变量，边代表变量之间的依赖关系。

如图 1 所示，$X=\{X_1,\dots,X_n\}$ 代表随机变量，若在图中存在箭头$X_i \rightarrow X_j$，i,j∈{1,2⋯n}，则表明 X_i 是 X_j 的父节点。根据马尔可夫假设，X_1、X_2 ... X_n的联合概率分布可以用式（1）表示。

$$P(X_1,\dots,X_n)=\prod_{i=1}^{n}P(X_i|\pi(X_i)) \tag{1}$$

其中：$\pi(X_i)$ 表示 X_i 的所有父节点。

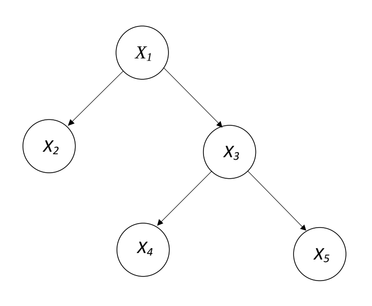


图 1　一个贝叶斯网络

1.2　因果理论

传统的统计学用相关性来描述随机变量间的关系。但是相关性不能完全概括随机变量之间的关系。例如可以很容易得到“刮风—树叶摇晃”的关联关系。根据常识，刮风是树叶摇晃的原因，有刮风→树叶摇晃。但反过来，则会认为树叶摇晃是刮风的原因，这显然是错误的。在关联关系中，存在着因果关系。

Pearl 对于这一系列问题，提出了因果理论。他对于事物间的因果性的强度用平均因果效应来描述，更加形式化的，用式（1）描述两事物 *X*、*Y* 之间的平均因果效应（average cause effective，ACE）。

$$\begin{aligned}ACE &= P(Y=1\,|\,do(X=1)) \\ &-P(Y=1\,|\,do(X=0))\end{aligned} \tag{2}$$

其中：$do(X=1)$ 为 Pearl 定义的“do—操作”。式（3）给出了“do—操作”的形式化定义：

$$\begin{aligned}&P(Y=y\,|\,do(X=x)) \\ &=\sum_{Z}P(Y=y|X=x,Z=z)P(Z=z)\end{aligned} \tag{3}$$

其中：*Z* 为 *X* 到 *Y* 的后门路径集合。

2　贝叶斯网络结构学习方法

由于 K2 算法需要为之提供节点优先次序，而节点优先次序只能定性地描述变量之间的父子关系，所以基于因果效应的方法第一步定义了节点优先度，用于定量地描述节点之间父子关系。2.1 节定义了节点优先度的计算方法，将节点优先度向量降序排列，得到节点优先次序。第二步通过节点优先次序，结合 K2 算法初始化贝叶斯网络。为提高运算效率，提出的节点优先度计算方法不考虑两节点之间的后门路径，虽然在不考虑后门路径的情况下，大部分节点优先次序已经是真实的优先次序，但部分节点优先次序和真实优先次序有一定偏差，得到的初始化网络存在反向边和多边的情况。所以第三步通过逐一将已有边反向和删除，寻找比初始化网络评分更高的网络代替初始化网络。为避免使用评分删除和反向调节时陷入局部最优，使用互信息删除边策略同时删除部分多余边，这样可以使评分从另一个新的起始点开始。最后，直到评分不再上升时结束算法。

2.1　节点优先度和节点优先次序

节点优先次序包括两个算法，分别从改进的因果效应和无边 BDe 评分出发，通过两个算法得到节点优先度向量。降序排

列该向量，得到节点优先次序。更加形式化地，有如下定义：

定义 1　节点优先度。对于任意的节点 X_i 、 $X_j \in V\{X_1,\ldots,X_n\}$ ，V 是贝叶斯网络中的节点集，n 为贝叶斯网络节点个数，对于节点集 $Y=\{X_1、X_2...X_{i-1}，X_{i+1}...X_n\}$ 和节点 X_i 。给定一个判断标准 S，存在 N 个节点 $\{x_k^1、x_k^2、x_k^N\}$， $X_k \in Y$，使得判断标准 S 成立，则 N 为节点 X_i 的优先度。

2.1.1 基于因果效应的节点优先度算法

使用如下形式近似数据集中任意两节点 X_i, X_j 之间的因果效应（cause effective） $CE_{X_i \to X_j}$ ：

$$CE_{X_i \to X_j} = \frac{N(X_j=1)}{N} * \left[P(X_j=1|X_i=1) - P(X_j=1|X_i=0)\right] - \frac{N(X_j=0)}{N}\left[P(X_j=0|X_i=1) - P(X_j=0|X_i=0)\right] \quad (4)$$

其中： $N(X_j=1)$ 表示 $X_j=1$ 的样本个数，N 表示总样本量。 $X_i=1$ 表示随机变量 $X_i=ture$， $X_i=0$ 表示随机变量 $X_j=false$ 。不考虑“do-操作”，并拓展 J.pearl 的因果效应。式（2）的 ACE 只考虑了 X 节点对于 Y 节点在 Y=1 处的因果效应。如式（4）所示，考虑了 $X_i \to X_j$ 在 $X_j=1$ 和 $X_j=0$ 两处的因果效应。

使用因果效应作为判断标准，算法从网络中的一个节点出发，依次计算该节点对于其他节点的因果效应，若 $CE_{X_i \to X_j} > CE_{X_j \to X_i}$ ，则 X_i 的节点优先度加一，否则 X_j 的节点优先度加一。算法遍历网络中任意两节点之间的因果效应。假设有 N 个节点，共进行$\sum_{x=1}^{N-1} x$次计算。降序排列节点优先度向量，得到节点优先次序。

2.1.2 基于 BDe 评分函数的节点优先度方法

BDe 是最早的贝叶斯网络数据拟合程度的评分函数之一，它假设数据服从狄里克雷分布（dirichlet distribution），有如下形式：

$$score(G|D) = \sum_{i=1}^{n}\sum_{j=1}^{q_i}\left[log\frac{\Gamma(\alpha_{ij*})}{\Gamma(\alpha_{ij*}+m_{ij*})} + \sum_{k=1}^{r_i} log\frac{\Gamma(\alpha_{ijk}+m_{ijk})}{\Gamma(\alpha_{ijk})}\right] \quad (5)$$

其中：Γ(*) 为伽马函数，m_{ijk} 表示数据中第 i 个节点，取第 k 个值，其父节点为第 j 个值的样本个数，α_{ijk}是狄利克雷分布的超参数，角标含义同 m_{ijk} 。 $m_{ij*}=\sum_k m_{ijk}$， $\alpha_{ij*}=\sum_k \alpha_{ijk}$ 。

使用 BDe 函数作为判断节点优先度的标准，得到基于 BDe 评分函数的节点优先度向量。算法开始从一个只有节点的网络出发。对于网络中所有节点 $X_i \in \{X_1,\ldots,X_n\}$ 和 $X_j \in \{X_1,X_2\ldots X_{i-1},X_{i+1}\ldots X_n\}$ 构建两个图 G，其中： G_1 的节点 X_i 指向 X_j， G_2 的 X_j 指向 X_i 。然后根据 BDe 评分函数计算 $score(G_1|D)$ 和 $score(G_2|D)$ ，若 $score(G_1|D)$ > $score(G_2|D)$，则 X_i 的优先度加一，反之亦然。降序排列节点优先度向量得到节点优先次序。

需要特别说明的是，若在节点优先度向量中，X 节点和 Y 节点的节点优先度相同，则使用当前节点优先度评判标准，仅针对 X 和 Y 两个节点判断节点优先顺序。

2.2　初始化网络

运用得到的节点优先次序结合 K2 算法学习初始化网络。在节点优先度计算环节中，大部分情况可以直接得到正确的节点优先次序，但是由于基于因果效应的节点优先度计算方法删除了“do-操作”和后门路径，基于 BDe 评分的节点优先度计算方法只考虑无边网络评分，所以节点优先度向量和真实优先度向量可能存在偏差，到这里得到的是一个近似正确的网络。

K2（cooper and Herskovits,1992）是贝叶斯网络结构学习算法之一，设随机变量 $D=\{X_1,\ldots,X_n\}$ 是一组完备数据集。K2 算法使用 CH 评分作为评分准侧，其形式如下：

$$scoreK2 = \sum_{i=1}^{n}\left[\sum_{j=1}^{q_i}\left[log\frac{(r_i-1)!}{(m_{ij*}+r_i-1)!} + \sum_{k=1}^{r_i} log(m_{ij*})\right]\right] \quad (6)$$

其中： m_{ijk} 表示数据中第 i 个节点，取第 k 个值，其父节点为第 j 个值的样本个数， $m_{ij*}=\sum_k m_{ijk}$ 。为了缩小搜索空间，需要给 K2 算法提供节点次序和最大父节点个数。排在节点次序前的随机变量不能成为排在节点次序后的随机变量的子节点。例如，若有节点顺序 $order=\{\ldots X_i,\ldots,X_j\ldots\}$，则 X_i 不可能是 X_j 的子节点。K2 算法的结果正确性很大程度取决于节点次序。节点次序一般通过先验知识获得，但在现实中，基本无法通过先验知识获得节点次序，以下提供了从数据中学习节点次序的方法。

2.3　反向调节和基于互信息的删边策略

2.3.1 基于互信息的删除边策略

大部分情况下，初始化的网络就是真实的贝叶斯网络。但也有可能由于节点次序的略微偏差导致网络中少量边是错误的。因此需要对这些边进行反向调整和删除操作。由于 2.3.2 的评分调节策略是在某一特定情况下，对网络的某些边进行反向和删除操作，容易陷入局部最优，无法更贴近全局最优。为了使评分的初始值更贴近全局最优，减少容易陷入局部最优的情况，使用互信息删除关联度较小的边(这些边在真实网络中不存在，在当前状态下的评分有可能高于删除后的评分)，从而更容易使 2.3.2 的评分方法贴近全局最优。互信息可以用来描述变量之间相关程度，用下式描述：

$$I(X;Y) = H(X) - H(X|Y) \quad (7)$$

其中：$H(X)$为变量 X 的信息熵，$H(X|Y)$为变量 X 对于变量 Y 的条件信息熵。变量 X 的信息熵可以用下式描述：

$$H(X) = -\sum_{X=x} P(x) * logP(x) \quad (8)$$

同样地，变量 X 和变量 Y 的条件信息熵可以用式（9）描述。

$$H(X|Y) = -\sum_{X=x}\sum_{Y=y} P(x|y) * logP(x|y) \tag{9}$$

信息熵描述了变量所含信息不确定性的多少。使用互信息代替条件独立测试，可以有效地提升计算效率。删除边策略是将所有节点之间的互信息进行降序排列，对于网络中的所有节点，依次作为目标节点 *T*，选择目标节点 *T* 的互信息排名靠前的 h 个节点作为可能和目标节点 *T* 有边的节点集。然后在初始化网络中，对于和目标节点有边连接的节点 Y 依次查看 a）*T* 节点是否在 *Y* 节点的前 h 个互信息倒序排列列表中;b）*Y* 节点是否在 *T* 节点的前 h 个降序排列的互信息节点列表中。若 a）b）都不满足，则删除目标节点 *T* 和 *Y* 节点之间的边。为了避免错误的删除正确边以及不能删除多余边两种情况，根据节点数量来选择 h 的数值。在实验部分，取 h=$[\frac{3}{8}n]$，其中 n 是节点个数。

2.3.2 基于 BDe 评分的反向调节和删边调节

完成了 2.3.1 节的环节后，已经将网络中大部分相关度不高的节点对之间的边删除，此时网络评分更接近全局最优。这时还需要调节部分由于节点次序错误导致的反向边。

在这个环节，使用 BDe 评分作为调节标准，使用反向边和删除边两个算子，对于网络中的边集 $E = \{e_1, e_2 \dots e_n\}$，依次将边 e_i 反向和删除，然后计算 BDe 评分，若 BDe 评分高于反向前网络的 BDe 评分，将边 e_i 反向或删除。直到评分不再增加停止算法。

3　实验结果

基于因果效应的方法在两个贝叶斯标准数据库下测试（ASIA 数据库、ALARM 数据库），并与 MMHC 算法、MCMC 算法、爬山法和随机 K2 算法进行对比。对于同一组数据，从正确边、丢失边、反向边、多边的平均值以及标准差进行对比。取不同样本量数据进行实验，样本量分别为(2000、4000、6000、8000、10000、12000、14000、16000、20000）.实验表明，该方法在正确边个数、错误边个数均胜过其他贝叶斯网络学习方法。

3.1　ASIA 数据库

ASIA 数据库是贝叶斯标准数据库之一，描述了胸腔诊断网络，如图 2 所示。实验结果如表 1 和图 3、4 所示。

可以看到，基于因果效应的方法在 10 组实验中，平均正确边数高达 7.0 条，而 MMHC 算法的平均正确边数仅为 6.3 条，正确边数量较 MMHC 算法提升了 11.1%。图 3 展示了不同算法在不同样本量下的正确边数量。在 ASIA 数据库中，MMHC 算法是对比算法中性能最为优越的。图 4 针对 MMHC 算法进行了正确边增量对比。为便于阅读，隐藏横坐标相应样本量，横坐标数值同图 3。

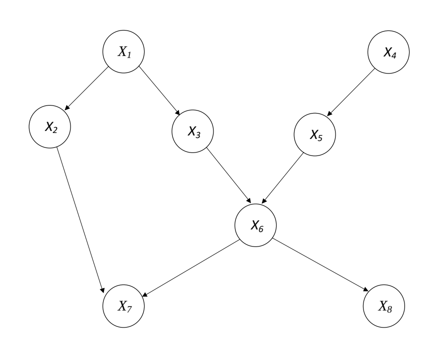


图 2　ASIA 网络

表 1　ASIA 网络学习结果比较分析

	基于因果效应的方法		爬山法		随机节点顺序的 K2 方法		MMHC 算法		MCMC 算法	
	平均	标准差	平均	标准差	平均	标准差	平均	标准差	平均	标准差
正确边	7.0	0.94	3.75	1.75	3.41	0.16	6.3	0.82	3.9	2.18
丢失边	0.1	0.32	0.96	0.65	1.32	0.36	0.5	0.71	0.5	0.71
反向边	0.8	0.79	3.29	1.75	3.26	0.26	1.2	0.63	3.5	1.78
多余边	0.7	1.06	1.68	1.31	4.67	0.24	0.6	0.84	4.0	1.70
图错误	1.5	1.51	5.93	2.81	9.27	0.28	2.3	1.26	8.0	3.50

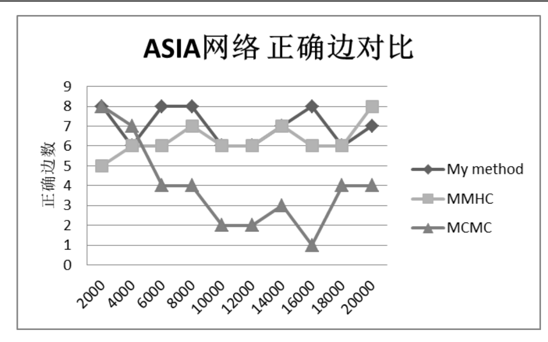


图 3　ASIA 网络正确边对比

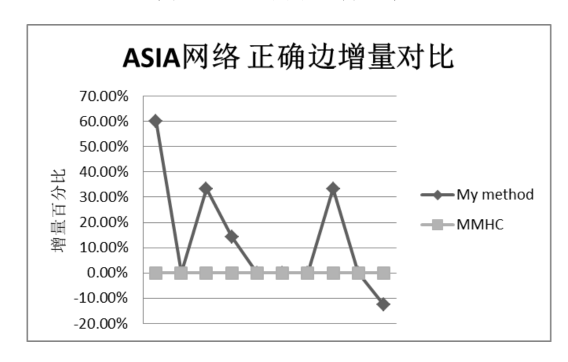
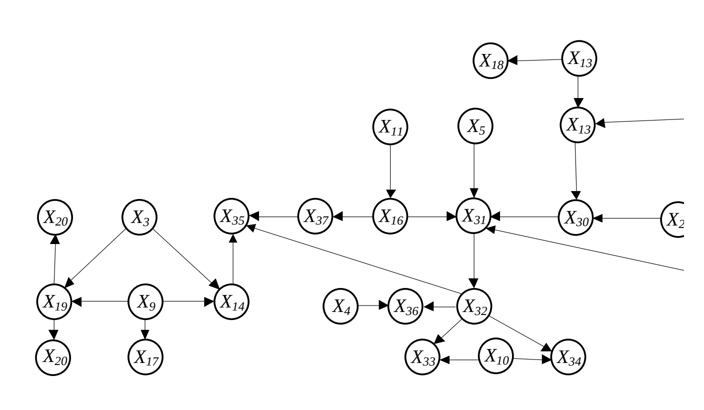


图 4　ASIA 网络正确边增量对比

实验表明，在 ASIA 数据集上，基于因果效应的方法比 MMHC 算法有更高的准确性，在十组实验数据中，有四组达到了全局最优，学习到了全部正确的网络结构（正确边为 8 条的网络结构），而 MMHC 算法仅有一组学习到了完全正确的网络结构。对于大多数样本来说，该方法较 MMHC 算法更容易达到全局最优。

3.2　ALARM 网络

ALARM 网络来源于医疗诊断监控系统，由 37 个节点和 46 条边组成。其网络结构如图 5 所示。


图 5 ALARM 网络结构

这里选用基于 BDe 评分的节点优先次序计算方法来进行实验。实验表明，该方法能准确学习贝叶斯网络。表 2 给出了实验结果。基于因果效应的方法在 ALARM 十组数据集中平均正确边数为 41.2 条，而 MMHC 算法的平均正确边数仅有 34.1 条，较 MMHC 算法正确率提升 20.8%。结合 ASIA 网络的实验结果，在 20 组实验数据中，平均正确率较 MMHC 方法提升 16%。

图 6 给出了不同算法在不同样本量数据下的正确边对比，MMHC 算法性能优于其他对比算法。图 7 给出了基于因果效应的方法较 MMHC 算法的增量对比。为方便对比，图 7 隐藏了横坐标数值，具体数值同图 6。

表 2 ALARM 网络学习结果比较分析

	基于因果效应的方法		爬山法		随机节点顺序的 K2 方法		MMHC 算法		MCMC 算法	
	平均	标准差	平均	标准差	平均	标准差	平均	标准差	平均	标准差
正确边	41.2	3.08	21	4.24	21.9	0.61	34.1	6.05	20	3.56
丢失边	0.5	0.53	2.5	0.71	2.40	0.25	0.3	0.48	3.5	2.11
反向边	4.2	2.53	22.5	4.94	21.8	0.54	11.6	6.19	22.5	2.83
多余边	5.5	2.32	12	2.82	40.8	1.45	4.1	1.45	44.5	8.03
图错误	10.2	4.80	37	7.07	64.9	1.66	16.0	6.93	70.5	10.54

可以看到，基于因果效应的方法在 ALARM 图的十个数据库中有 8 个数据库好于 MMHC 算法，在丢失边和反向边上都好于 MMHC 算法。由于使用 K2 算法得到初始化网络，当节点优先次序错误时，会导致多余边数量的上升。在第二部分删除边调节中，仅考虑同时最多删除一条边的评分比较，这虽然可以提高运算效率，但是还是存在陷入局部最优的可能性，从而导致多余边在某些情况下不能完全删除。图 8 给出了两个网络，20 组实验结果的平均标准差数据对比。

基于因果效应的方法，正确边标准差平均较 MMHC 低 17%。其中 ASIA 网络中，正确边标准差无明显差异，这是因为 ASIA 只有 8 个节点 8 条边，相对于大型网络，算法学习难度较低，不容易陷入局部最优。而在 alarm 网络中，由于共有 46 条边需要学习，学习难度较大，容易陷入局部最优，所以标准差差距也相对较大。对比随机 K2 方法的平均标准差，虽然随机 K2 算法的标准差性能是最好的，但是如图 9 所示随机 K2 方法准确率明显低于基于因果效应的方法。分析随机 K2 方法的图结果发现，当节点顺序有较大偏差时，虽然正确边和爬山法、MCMC 相差不大，但随机 K2 方法生成的网络更偏向于全连通网络，应用价值非常低。图 9 给出了算法的正确率对比。可以看到，在两个网络的平均正确率上，该方法较 MMHC 有明显的提升。

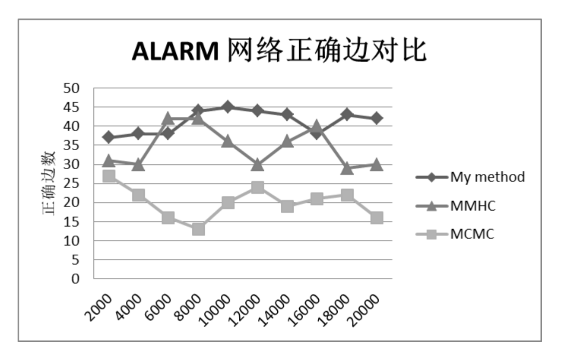


图 6 ALARM 网络正确边对比

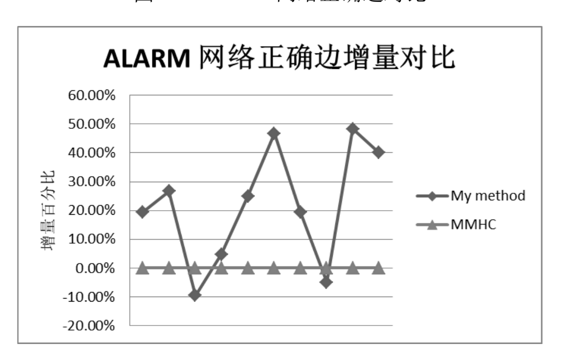


图 7 ALARM 网络增量对比

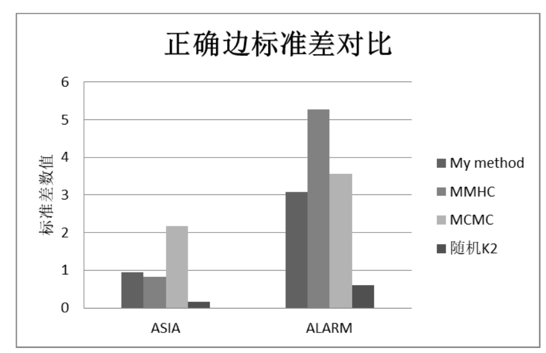


图 8 不同方法标准差对比

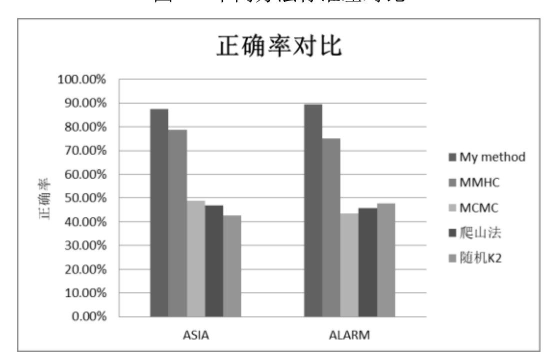


图 9 不同方法正确率对比

实验表明，基于因果效应的方法较其他算法有更好的准确性和稳定性，在 ASIA 数据库上胜过 MMHC 算法，远超爬山法、随机 K2 法和 MCMC 算法。在大型网络 ALARM 中，随着样本数的增加，较 MMHC 算法的效率也随之提升，远超过 MMHC、MCMC 等算法，有更好的准确性和稳定性。

4　结束语

贝叶斯网络是用来描述节点联合概率的图模型，贝叶斯网络中的箭头可以表示因果性。提出了基于 pearl 因果效应和基于 BDe 评分函数的节点优先度算法，结合 K2 算法在贝叶斯网络标准数据库上胜过 MMHC 等算法。基于因果效应的方法在调节反向边环节，是通过评分函数调节的。目前已有的评分函数，都需要假设数据服从某一个先验分布。这使得评分函数不能准确的描述数据和网络的拟合程度。考虑到在今后的研究中，可以通过文本挖掘关键词网络中有相关关键词的句子，用自然语言处理的方法识别出贝叶斯网络箭头的方向，从而调整从数据中学习得到的图的部分错误边方向。这也是今后贝叶斯网络学习的发展趋势。

参考文献：

[1] Seixas F L, Zadrozny B, Laks J, *et al.* A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer′s disease and mild cognitive impairment [J]. Computers in Biology and Medicine, 2014, 51 (7): 140-158.

[2] Ramazzotti D, Graudenzi A, Antoniotti M. Modeling cumulative biological phenomena with Suppes-Bayes causal networks: [J/OL]. Evolutionary Bioinformatics Online, 2018, 14. (2016-02-25) [2018-02-26]. http://dx. doi. org/10. 1101/041343.

[3] Fenton N, Neil M, Marquez D. Using Bayesian networks to predict software defects and reliability [J]. Journal of Risk & Reliability, 2017, 222 (222): 701-712.

[4] 金杉，崔文，金志刚．正态分布的贝叶斯网络火灾数据融合预警研究 [J]．计算机应用研究, 2016, 33 (5): 1473-1476.

[5] Sein M，傅顺开，吕天依，等．一般贝叶斯网络分类器及其学习算法 [J]．计算机应用研究, 2016, 33 (5): 1327-1334.

[6] He Lianghua, Hyu Die, Wan Meng, *et al.* Common Bayesian network for classification of EEG-Based multiclass motor imagery BCI [J]. IEEE Trans on Systems Man & Cybernetics Systems, 2017, 46 (6): 843-854.

[7] 高瑞，王双成，杜瑞杰．企业运行指标因果分析的动态贝叶斯网络方法 [J]．计算机应用研究, 2016, 33 (5): 1433-1436.

[8] Kaiser J L, Bland C L, Ii D J K. Identifying causal networks linking cancer processes and anti-tumor immunity using Bayesian network inference and metagene constructs [J]. Biotechnol Prog, 2016, 32 (2): 470-479.

[9] Klinger T, Rottensteiner F, Heipke C. A dynamic Bayes network for visual pedestrian tracking [J]. ISPRS-International Archives of the Photogrammetry: Remote Sensing and Spatial Information Sciences, 2014, XL-3 (3): 145-150.

[10] 徐建民，唐万生，陈振亚．贝叶斯网络在信息检索中的应用 [J]．河北大学学报：自然科学版, 2007, 27 (1): 93-98.

[11] Chickering D M, Heckerman D, Meek C. Large-sample learning of Bayesian networks is NP-Hard [M]. brookline: JMLR. org, 2004.

[12] Zhao bo, Wyu Qingchang, Yin Shitang, *et al.* An improved Bayesian network structure learning algorithm based on the conditional independence test [J]. Journal of Yunnan University of Nationalities, 2011, 20 (5) .

[13] Geng Zhi, Wang Chi, Zhao Qiang. Decomposition of search for v-structures in DAGs [J]. Journal of Multivariate Analysis, 2005, 96 (2): 282-294.

[14] Lam W, Bacchus F. Learning Bayesian belief networks: an approach based on the MDL principle [J]. Computational Intelligence, 1994, 10 (3): 269-293.

[15] Cooper G F, Herskovits E. A Bayesian method for the induction of probabilistic networks from data [J]. Machine Learning, 1992, 9 (4): 309-347.

[16] Li Guoling, Xing Lining, Zhang Zhongshan, *et al.* A new Bayesian network structure learning algorithm mechanism based on the decomposability of scoring functions [J]. IEICE Trans on Fundamentals of Electronics Communications & Computer Sciences, 2017, 100 (7): 1541-1551.

[17] Chen Xuewen, Anantha G, Lin Xiantong. Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm [J]. IEEE Trans on Knowledge & Data Engineering, 2008, 20 (5): 628-640.

[18] 张连文．贝叶斯网引论 [M]．北京：科学出版社, 2006.

[19] Acid S, Campos L M D. Searching for Bayesian network structures in the space of restricted acyclic partially directed graphs [M]. USC: AI Access Foundation, 2003.

[20] Liu Xuqing, Liu Xinsheng. Structure learning of Bayesian networks by continuous particle swarm optimization algorithms [J]. Journal of Statistical Computation & Simulation, 2018, 88 (9): 1-29.

[21] Lerner B, Malka R. Investigation of the K2 algorithm in learning bayesian network classifiers. [J]. Applied Artificial Intelligence, 2011, 25 (1): 74-96.

[22] Cooper G F, Herskovits E. A Bayesian method for the induction of probabilistic networks from data [J]. Machine Learning, 1992, 9 (4): 309-347.

[23] Tsamardinos L, Brown L E, Aliferis C F, *et al.* The max-min hill-climbing Bayesian network structure learning algorithm. [J]. Machine Learning, 2006, 65 (1): 31-78.

[24] Pearl J, Glymour M, Jewell N P. Causal inference in statistics: a primer [M] , Hoboken: Wiley, 2016: 53-87.